图书情报工作
第 65 卷 第 1 期　2021 年 1 月

国际开放科学研究进展*

■ 黄如花　赵洋　黄雨婷
武汉大学信息管理学院　武汉 430072

摘　要：[目的/意义]开放科学可以使科学研究更具协作性、透明性和高效性，已引起全世界的广泛关注。从国际视角梳理和分析开放科学当前研究主题有利于推动该领域发展，为后续研究提供参考和借鉴。[方法/过程]运用文献调查与信息可视化相结合的方法，选取 Web of Science 平台收录的关于开放科学研究的论文，并结合相关政府文件、研究报告、新闻报道等文献，揭示国际开放科学研究的主题分布。[结果/结论]国际开放科学的研究主题包括：开放获取研究、数据共享研究、成果重用研究、知识创新研究、基础设施建设研究。未来建议从政策体系、基础设施、主体参与 3 个方面进行系统、深入研究。

关键词：开放科学　开放数据　开放获取　数据共享
分类号：G250.73
DOI：10.13266/j.issn.0252－3116.2021.01.019

开放科学将带来科学研究的重大革命，已引起全球科学界的广泛关注。2020 年 3 月，为探讨新型冠状病毒疫情背景下国际科学合作的重要性，联合国教科文组织召集多国科学部门代表举行线上会议，呼吁加强"开放科学与合作"[1]。此前，教科文组织已于 2019 年 9 月启动制定《联合国教科文组织开放科学建议书》（*The UNESCO Recommendation on Open Science*）[2]。从学术研究来看，国外学者围绕开放科学的概念[3]、框架[4]、实践[5-6]等方面开展了较为全面的研究，而我国的研究还处于概念辨析[7-8]、经验借鉴[9]的起步阶段，尚未建立起完善的开放科学研究体系。

基于此，本文拟采用定量与定性相结合的方法，对国际开放科学研究情况进行述评，力图梳理该领域的研究进展，为后续相关研究提供参考。

1　研究设计

开放科学作为一个新兴主题，其内涵与外延等基本问题尚未达成共识。因此，在进行文献分析之前，有必要对开放科学的概念进行阐释，并据此确定检索策略。

1.1　开放科学的概念

目前学界和业界尚未就开放科学的概念达成一致见解。开放科学研究的五大学派包括基础设施学派（涉及技术架构）、公共学派（涉及知识创造的可及性）、计量学派（涉及替代计量学）、民主学派（涉及知识获取）和务实学派（涉及合作研究）[10]。本文主要参考欧盟委员会（以下简称"欧盟"）对开放科学的定义，即通过数字化工具和网络实施、传播并转变科学研究的方式，开放科学取决于技术发展和文化变革对科研合作和开放的共同影响。通过提供无限、无障碍、开放获取的研究成果，开放科学旨在使科学研究过程更加高效、透明和有效[11]。由欧盟资助开发的 FOSTER portal 电子学习门户，构建了开放科学分类体系，将开放科学按主题领域划分为开放获取、开放数据、开放可重复研究、开放科学评估、开放科学指南、开放科学政策、开放科学项目、开放科学工具等[12]。

1.2　数据来源与研究方法

本文数据主要来源于 Web of Science 核心合集数据库。考虑到围绕开放科学不同构成要素（如开放获取、开放数据、开放资源等）的研究较多，若检索词涵盖所有构成要素，则检索结果非常庞大，且无法从宏观层面研究开放科学。因此，本文将检索式设为 TS =（"open science" OR "opening science"），选择 SCI-E、SSCI、A&HCI 三大引文索引，文献类型限定为 article，时间范围为 1900－2019 年，共检索得到 731 条数据。手动剔除不相关数据后，得到有效数据 386 条，包含全

作者简介：黄如花（ORCID：0000－0003－4250－1914），教授，博士生导师，博士；赵洋（ORCID：0000－0003－1705－1069），博士研究生，通讯作者，E-mail：zhaoyang_whu@163.com；黄雨婷（ORCID：0000－0003－2444－4856），博士研究生。
收稿日期：2020－04－30　修回日期：2020－09－18　本文起止页码：140－149　本文责任编辑：王传清

黄如花，赵洋，黄雨婷. 国际开放科学研究进展[J]. 图书情报工作,2021,65(1):140－149.

记录与引用的参考文献。从发表年份来看，开放科学研究始于 1985 年，2014 年起论文数量快速增长，这与近年各国和国际组织的政策引导有着密切关系。为更加全面地了解国际开放科学发展的最新动态，本研究以 386 篇文献为主，兼顾相关政策文件、研究报告、研究计划、新闻报道等文献。

本文选用 VOSviewer 可视化分析软件，该软件可构建基于关键词的共现网络，聚类分析效果较好。关键词凝练了一篇文章的精华，可以在一定程度上反映文章的核心思想。本文将关键词频次阈值设定为 2，通过对关键词进行预处理，包括将单复数、大小写、简称或全称的词汇形式统一描述，将语义上表达一致的同义词、近义词归并为通用词，删除无实际意义、与主题不相关的词汇等，最终选取 332 个关键词绘制开放科学研究关键词共现聚类图谱。

如图 1 所示，共有 5 个主要聚类，聚类之间相互交叉和关联，联系紧密。图 1 中节点的衡量指标包括出现频次（Occurrences）和总联系强度（Total link strength），分别代表关键词出现频次和与其他关键词的总共现次数，据此可以判定核心关键词，表 1 展示了排名前 20 的核心关键词。下文将按照聚类大小，结合核心关键词和文献内容，依次对 5 个研究主题展开分析。

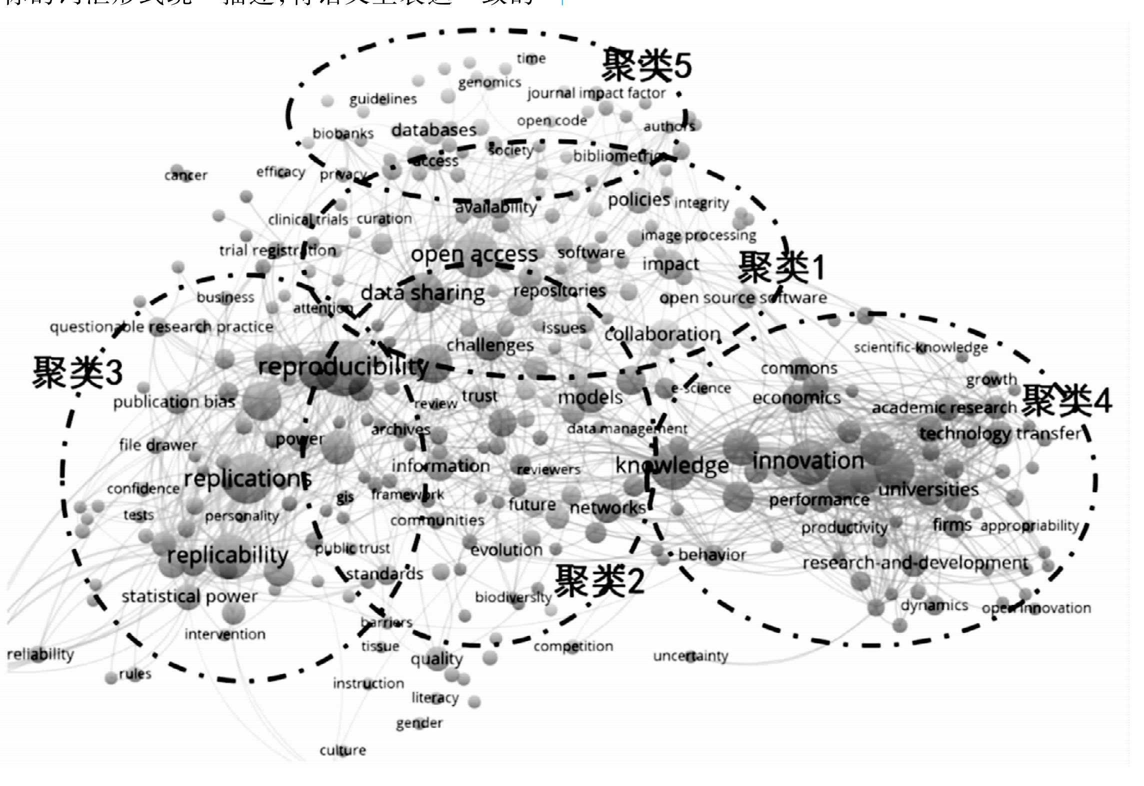


图 1　国际开放科学研究关键词共现聚类

表 1　国际开放科学研究核心关键词统计(按出现频次排序)

关键词	出现频次	总联系强度	关键词	出现频次	总联系强度
reproducibility	47	74	open data	21	33
replications	38	69	technologies	19	37
open access	31	25	patents	16	35
knowledge	30	45	incentives	16	30
innovation	29	46	industry	15	28
replicability	28	53	universities	14	26
intellectual property	23	36	preregistration	13	31
transparency	22	31	meta-analysis	13	26
data sharing	22	27	collaboration	13	21
psychology	21	40	models	13	12

2 关于开放获取的研究

聚类1是最大也是最核心的聚类,关键词节点分布较为密集,形成了以开放获取(Open Access,OA)为核心的研究主题,其他核心关键词还包括模式、影响、网络、信息、系统等。开放获取自20世纪90年代被提出以来,一直是学术界和业界的热门话题。2002年发布的“布达佩斯开放获取先导计划”(Budapest Open Access Initiative,BOAI)中,首次明确了开放获取的定义[13]。随后,开放获取运动进入蓬勃发展阶段,被认为是开放科学的重要实践[14]。

2.1 开放获取的模式

学界广泛地讨论了开放获取的各种模式,以提高开放获取效果,如按出版载体和可用程度划分的多种出版模式,以开放同行评审(Open Peer Review)为主的评审模式等。

2.1.1 开放获取的出版模式

开放获取的出版模式如表2所示,其中,绿色OA(Green OA)和金色OA(Gold OA)是最主要的两种出版模式。M. Laakso等根据期刊内容的可用性,将金色OA细分为直接OA(Direct OA)、延迟OA(Delayed OA)和混合OA(Hybrid OA)[15]。也有学者将混合OA与绿色OA、金色OA并列,认为三者同为开放获取出版模式的主要类型[16]。此外,较为常见但讨论不足的出版模式还包括棕色OA(Bronze OA)[16]、黑色OA(Black OA)[17]等。

表2 开放获取的出版模式

OA类型	相关概念
绿色OA	作者将已出版或未出版的文献存储到机构库、知识库等平台供大众免费获取
金色OA	作者将文章发表在开放获取期刊上供读者免费获取和阅读
混合OA	作者向学术出版商付费以使期刊中的文章可自由访问
棕色OA	在没有明确的开放许可情况下,就可以在出版商网站上免费阅读的文章
黑色OA	在Sci-Hub和LibGen等非法网站上共享的文章

很多学者对不同的开放获取模式进行了调查和比较。J. C. Valderrama-Zuri[18]、L. Zhang[19]分别从教育、物理等研究领域比较了不同的开放获取模式。H. Piwowar等通过开放在线服务工具oaDOI,发现受金色OA和混合OA的推动,开放获取文章的比例正在增加[16]。通过研究开放获取文章的引文影响,H. Piwowar等还证实了在绿色OA和混合OA影响下,开放获取文章获得的引文量比平均水平高18%[16]。

2.1.2 开放获取的评审模式

当前,开放获取模式导致了掠夺性期刊(Predatory Journals)的出现,即以非常粗浅的方式进行同行评审,并在几天内录用稿件以收取出版费用[20]。而开放同行评审不仅可以遏制掠夺性期刊,还可以形成更加严格的(通过与学术团体之间的交流)和相关的(通过与其他利益相关者的交流)研究[20]。T. Ross-Hellauer对开放同行评审进行了系统性的述评,将开放同行评审归纳为公开评审者和作者身份、发布评审报告以及使更多专家参与评审[21]。L. Dobusch和M. Heimstädt通过分析管理研究中的掠夺性出版,呼吁实行开放同行评审,认为其有利于实现评审过程的透明化,区分高质量期刊与掠夺性期刊,建立可持续发展的评审文化[20]。

2.2 开放获取的影响

开放获取作为一种新型出版模式,有利于推动科研成果通过网络自由传播,促进学术信息的交流与利用,对学术、经济和社会都产生了重大影响。

2.2.1 开放获取的学术影响

开放获取的学术影响主要体现在学术引用和学术交流方面。从学术引用来看,开放获取的文献通常更具有引用优势[22-23],虽然这种引用优势在不同的学科和研究领域表现不同,但研究的总体趋势表明,在引文方面,开放获取文献有利于提高学术影响力[24]。从学术交流来看,开放获取既能增加学者对学术资源的可用程度,又能促进他们研究成果的扩散[25]。X. Wang等认为开放获取文献在替代计量学方面也更具优势[26],更容易获得社交媒体关注,从而在非正式学术交流中产生更广泛的学术影响。

2.2.2 开放获取的经济影响

开放获取的经济影响体现在经费和成本两方面。在经费方面,开放获取不仅可以为出版商和研究资助者节省财政支出,还有利于提高传统订阅型期刊的经济效益[27]。R. P. Holley指出,大型商业出版商可以通过收取论文处理费(Article Processing Charges,APCs)和收购小型出版社从开放获取中受益[25]。从成本来看,对研究结果和数据的开放获取可以节省访问成本、人工成本和交易成本。M. J. Fell认为如果开放获取允许更高的可访问性和结果使用率,那么研发(Research and Development,R&D)回报率会更高[28]。与付费期刊相比,开放获取期刊更具市场性、实用性和吸引力,将提供更大的访问量、可见性和忠诚度[6]。

2.2.3 开放获取的社会影响

开放获取不仅能使学者获益,还对社会产生了广泛影响。J. P. Tennant等提出,开放获取文献支持任

何可以上网并且能够检索和阅读材料的用户进行研究,为推动公众科学(Citizen Science)和终身学习提供了公平的环境[27]。S. E. Hampton 等认为开放获取可以支持资源贫乏的学者、机构和国家的学术资源获取,有利于进一步促进科学民主化和知识多样化[29]。

3 关于数据共享的研究

聚类 2 形成了以数据共享(Data Sharing)为核心的研究主题,核心关键词包括数据共享、透明度、开放数据、政策、挑战等。聚类 2 分布较为分散,与聚类 1 存在部分交叉,体现了开放和共享的紧密联系。数据共享不仅能提高研究的透明度和可重复性,促使研究人员改善研究质量,而且其他研究人员在使用共享数据时还可能激发二次创新,推动研究领域发展[30－31]。

3.1 数据共享政策

各国政府、国际组织、研究资助机构、出版商、学术期刊、高校、图书馆等主体都出台了包括数据共享在内的研究数据政策,以推动开放科学发展。国家层面来看,美国、澳大利亚、英国是发布政策较多的国家。欧盟在其新的研究与创新框架计划——“地平线 2020”(Horizon 2020)[32]中宣布要全面实施科研数据开放政策,以此进一步推动“开放科学”战略。A. Zuiderwijk 和 M. Janssen 通过建立开放数据政策的比较框架(包括环境和背景、政策内容、绩效指标和公共价值等因素),比较了荷兰不同政府级别的 7 项政策[33]。

研究资助机构、出版商和期刊的研究数据政策是促使研究人员共享研究数据的关键因素,对三者的政策研究是当前研究的焦点。C. Neylon 以多个国家和国际组织的研究资助机构为对象,分析研究数据共享(Research Data Sharing,RDS)政策的目的和模式,为资助机构提出政策建议[34]。L. Jones 等分析了 Taylor&Francis、Springer Nature、Elsevier、Wiley 等出版商提供的多层次数据共享政策[35]。由于大多数学科都通过期刊出版物来传播新的研究成果,因此期刊数据共享政策对研究人员何时何地提供研究数据具有很大的影响。多位学者从不同领域和学科分析了期刊数据共享政策的进展和模式[36－37]。D. M. Gorman 通过分析 250 种期刊的研究数据政策,指出当前各期刊的研究数据政策急需标准化和统一[38]。为制定研究数据政策的通用框架,由欧盟、美国国家科学基金会、美国国家标准与技术研究院以及澳大利亚创新部共同发起的研究社区组织——研究数据联盟(Research Data Alliance,RDA)于 2018 年发布《期刊和出版商研究数据政策总体框架》(*Journal and Publisher Research Data Policy Master Framework*)[39],确保期刊和出版商继续支持和鼓励跨学科研究的数据共享。

3.2 数据共享的挑战

虽然大多数研究人员对数据共享持积极态度[40－41],但普遍认为数据共享还存在诸多障碍和挑战,主要包括:①数据共享过程中涉及的隐私及伦理等法律、规范或道德问题[42－44]。例如在临床医学领域,数据通常涉及患者隐私,如果毫无限制地共享数据将引发一系列问题;关于机密数据的共享问题更是关系到国家安全。②数据共享的便利性和技术问题[30,45]。数据共享的便利性和技术支持如果达不到预期,可能会造成时间限制、资源限制和资金限制等问题。③数据共享导致的利益竞争。F. Rockhold 等指出有些研究人员可能会担心商业或学术竞争对手从基于共享数据的分析中获益[46]。也有学者从全球视角揭示了数据共享面临的挑战,包括个人层面时间不足和数据滥用;机构层面数据共享培训不够,缺乏激励机制和内部政策保障;国际层面薄弱的政策、道德和法律规范,缺乏数据基础架构以及互操作性问题[47]。

3.3 数据共享的实践

各学科领域开发了众多数据共享项目和平台,为数据共享实践提供借鉴和参考。耶鲁大学开放数据获取项目(The Yale University Open Data Access,YODA)为改善公共健康和促进数据共享开发了数据共享平台,给研究者访问临床试验数据提供了便利[48]。加拿大安大略省脑研究所(The Ontario Brain Institute,OBI)依靠其大规模信息平台 Brain-CODE 推动神经科学领域的研究数据共享,该平台稳健且可扩展的数据治理结构值得各数据共享平台借鉴[49]。化学信息学家也将其模型公开提供给学界,通过 Chembench 这一开放获取的网络信息门户,来自不同领域的研究人员可以进行化学基因组数据的管理、可视化、分析和建模[50]。

4 关于成果重用的研究

聚类 3 的核心关键词包括可再现性(Reproducibility)、复制、可复制性(Replicability)、激励、心理学、预注册等,形成了以可再现性和可复制性为中心的聚类,研究主题集中在以成果重用为目的的开放科学实践研究。

4.1 可再现性

可再现性是指使用与原始研究相同的数据、计算步骤、方法和代码以及分析条件获得一致的研究结

果[51]。当数据可用、可发现、可再现且描述充分时,科学家可以避免重复研究,并在已有的研究基础上进行创新[21]。

目前,生态学、天文学、气候学、神经科学和海洋学等领域都很重视研究成果的可再现性。如何提高研究的可再现性,成为学者关注的热点。B. Marwick 总结了可再现性研究的基本原则[52]:公开提供数据和代码并将其存储在适当的位置,使用编程语言编写用于数据分析和可视化的脚本,使用版本控制来管理文件的多个版本以及合作者的贡献,记录并共享用于数据分析的计算环境。J. J. Brito 等提出通过教授计算技能、开发数据和软件、实施可重复性研究、实行激励机制来提高生物医学领域的可再现性研究[53]。此外,建立基于科学工作流程的管理系统,利用结果可再现性工具,审计出版物可再现性也是提高可再现性研究的重要方法[51]。在图书馆领域,由莱顿大学图书馆数字学术中心组织的可再现性黑客马拉松(ReproHack NL)为如何通过社区参与提高研究的可再现性提供了建议和参考[54]。

4.2 可复制性

可复制性比可再现性要求更加严格,是指在解决同一科学问题的研究中,使用新数据或新的计算方法获得一致的研究结果[51]。与可再现性相比,可复制性要求必须执行实验并再次收集数据。由美国开放科学中心(Center for Open Science, COS)制定的《推进透明与开放指导方针》(*Transparency and Openness Promotion Guidelines*, *TOP Guidelines*)将可复制性作为透明和开放模块化标准之一,用 3 个层次等级划分期刊对于可复制性研究投稿的支持程度[55],体现了可复制性在开放科学实践中的重要性。

近年来,心理学、神经科学和其他相关领域频繁出现"复制危机"(Replication Crisis),原因包括研究能力不足、出版偏见、理论不精确、统计程序不完善、样本量不足、分析方法不得当等[56-57]。为提高研究的可复制性,M. J. Larson 建议降低研究人员在数据分析时的灵活性,报告样本量信息,通过合作增加样本量,完善报告标准(遵循已建立的报告数据指南,并增加对预注册和已注册报告的采用)[57]。其他建议还包括建立开放性准则,给予开放科学徽章和奖励,出版前检查错误和异常结果,鼓励发表全部研究成果,鼓励可复制的研究和制定有效的资助者政策[51]等。

5 关于知识创新的研究

聚类 4 相较于其他聚类,核心关键词的频次较高,包括知识、创新、知识产权、技术、专利等,体现了以知识创新为核心的研究主题,具体表现为对专利制度和开放许可(Open License)的研究。如何在知识产权保护和开放科学之间建立平衡,利用知识产权促进知识创新,成为学界关注的重点。

5.1 专利制度

专利制度如何适应开放科学的发展是当前研究的焦点。许多实证研究表明,专利制度在某种程度上阻碍了开放科学的发展[58]。科学家推迟了他们希望获得专利的研究成果发表时间,不愿意与其他科学家分享研究材料和成果。与预期目的相反,专利制度不一定会给专利申请人提供激励,促使其披露关于发明的信息。此外,J. S. Gans 等指出,专利质量可能与相关论文的科学影响程度呈负相关[59]。

值得注意的是,所有权本身并不是开放和共享的障碍。专利制度旨在公开有关发明的信息,明确所有权的归属是分享的必要前提[4]。相反,授权方式可能会对开放科学产生重要影响。一些研究机构,如蒙特利尔神经研究所(Montreal Neurological Institute),正在探索使用"无专利"和开放使用的方法推动科研成果开放共享[60]。他们认为这是开放科学的基本要求——除了规范引用要求外,用户可以不受限制地利用公共资助的科研成果。为解决专利制度与科学成果开放共享之间的矛盾,M. Bentwich 基于国际公认的临时专利申请(Provisional Patent Applications, PPA)提出强制性的临时专利论文申请(Provisional Patented Paper Applications, PPPAs)[58],要求专利申请人在保护专利和知识产权完整性的同时,为进行研究的他人使用其发明授予简易许可证,使专利申请人有义务通过科学出版物与他人共享科研成果。

5.2 开放许可

开放许可是指对用户使用的材料或产品施加很少限制的许可[61],研究对象主要涉及开放获取出版物、数据以及开源软件和硬件。

开放获取出版物通常采用创作共享(Creative Commons, CC)许可。CC 许可提供了 6 个基本的示范条款,其中 CC BY 和 CC BY-SA 为免费许可;另一类使用较多的内容开放许可是针对手册、教科书、其他参考资料和教学材料的 GNU 自由文档许可(GNU Free Documentation License, GFDL)。对数据而言,除了 CC 许可,开放数据共享网站(www.opendatacommons.org)还专门为数据库创建了一组许可协议。同时,英国已启动了"开放政府许可",有利于政府和其他公共部门的

数据重用[62]。开源软件的许可问题受到了相当多的关注[63],开放许可允许自由使用、修改和共享软件。对于开源硬件,开源硬件协会建议使用从 CC BY-SA 等一般许可到 CERN 开放硬件许可和 TAPR 开放硬件许可等更具体的许可[64]的 7 项许可。R. M. Luis Felipe 等描述了欧盟最新的开放许可协议,尤其是对开源硬件的规定,并阐述了将开源软件许可条款扩展到开源硬件的前景和挑战[65]。

6 关于基础设施建设的研究

聚类 5 的关键词节点较少,核心关键词包括合作、数据库、演变等,体现了以基础设施建设研究为核心的研究主题。基础设施在开放科学实践中的表现主要体现在:分布式计算,即利用多用户的计算能力进行研究;科学家的社会和协作网络,以支持研究人员的互动和协作[10]。电子基础设施(E-infrastructures)、开放科学工具、数据知识库(Data Repository)等基础设施的研究构成了当前的研究主题。

6.1 电子基础设施

启用电子基础设施可确保通过数据知识库提供的科学信息和数据进行交换和分析。它包括嵌入在一些知识库中的高级服务、超级计算设施和分布式计算网络[10]。如开放科学网格(Open Science Grid)通过将多台计算机连接到高性能计算机网络,支持大规模的数据密集型研究项目。M. Altunay 等指出,电子基础设施为科学家和研究人员提供了协作式研究环境,以共同解决分布式计算问题[66]。然而,电子基础设施会受到国家安全、隐私和保密、商业敏感性和知识产权等因素的影响;数据收集、提供和存储的成本问题,对国际数据流动的挑战也是电子基础设施面临的重要问题[4]。

6.2 开放科学工具

开放科学工具研究主要包括 3 类。第一类是研究全流程的工具,如 D. De Roure 等介绍的用于发现、分享和策划科学工作流程和实验计划的社交网站 myExperiment[67]。第二类是以集成方式满足多种需求的平台,如美国开放科学框架(Open Science Framwork)。第三类是研究各个阶段中的数字工具,如开放实验室笔记本、协同写作工具、开放工作流、协作书目、文本和数据挖掘类的共享工具等。随着各种工具的研发,研究人员需要花费更多时间学习和更新技能。如何提高开放科学工具的交互性和易用性,成为研究亟待解决的问题。

6.3 数据知识库

数据知识库在支持开放科学、与其他电子基础设施建立联系以及集成数字平台的研究数据管理方面发挥了关键作用[4]。一些在线存储库,如 runmycode. org,还提供了对研究出版物的代码和数据的公共访问。现有研究主要集中在数据知识库的数据共享与开放方面,如 K. Nishikawa 从知识共享的角度调查了日本数据知识库如何管理研究数据,构建了基于资源开放性、社区开放性和基础设施开放性的分析框架[68]。V. Xafis 和 M. K. Labude 概述了通过数据知识库进行数据共享的策略,并讨论了与健康相关的数据共享所引发的一系列道德问题[69]。数据知识库目前面临的问题主要包括国际共享标准不统一、跨国网络治理不规范[4]、商业模式如何可持续发展[70]等。

7 总结与展望

7.1 总结

国际开放科学研究涉及的领域非常广泛,涵盖心理学、教育学、信息科学、生物学、经济学等多个学科,呈现高度跨学科的特点。研究国家以欧洲和北美洲为主,主要包括美国、英国、德国、加拿大、法国、荷兰、意大利、西班牙等。核心研究机构包括伦敦大学学院、麦吉尔大学、牛津大学、北卡罗来纳大学教堂山分校、威斯康星大学等高校。研究主题集中在开放获取研究、数据共享研究、成果重用研究、知识创新研究和基础设施建设研究。

整体来看,国际开放科学的研究处于发展阶段,对子主题如开放获取、开放数据、数据共享等研究较多,而对开放科学宏观层面的研究偏少。研究不足在于:①缺乏政策内容体系的系统化研究。目前对开放科学政策的研究多集中在开放获取、数据共享政策及策略方面,研究较为具体化,而对开放科学宏观政策体系研究不足。②缺乏贯穿科研全流程的基础设施研究。目前的研究虽然涉及多个开放科学平台和工具,但研究点较为单一和零散,且位于主题聚类的最边缘,研究尚未成熟。③关于开放科学主体的研究深度不足。目前的研究虽然涉及研究人员、图书馆、出版商等主体,但对各主体如何参与开放科学实践的研究还存在不足,需进一步探讨如何提高各主体的参与意识和技能等。

7.2 未来展望

基于以上分析,并结合国内外政策与实践情况,建议未来从以下 3 个方面对开放科学进行系统、深入地

研究。

7.2.1 重视系统化的政策体系研究

开放科学政策能为实践提供指导和依据。相关研究可围绕国家层面的政策内容体系、面向多主体的政策执行机制、与知识产权等相关政策的协同发展策略开展。

(1)国家层面的政策内容体系。开放科学是联合国《2030 年可持续发展议程》(*The 2030 Agenda for Sustainable Development*)的核心推动力[71]。随着开放科学日益全球化以及不同层级的开放科学政策增加,如何构建国家层面的开放科学政策内容体系成为关注的重点。除了当前探讨较多的开放获取政策和开放数据政策,研究还应关注开放科学的其他构成要素,从科研全流程的视角进行思考,尤其要关注成果发表前的选题、思路和方法以及成果发表后的学术交流。

(2)面向多主体的政策执行机制。由于开放科学的发展涉及众多利益主体,如研究资助组织、研究执行机构、研究人员、出版商、图书馆等,既需要宏观的政策指导,还需要微观的政策实施细节[72]。研究应明确各利益主体的权利和义务,为制定面向各利益主体的专门性政策提供理论依据,如面向研究资助机构的科学研究资助政策、面向研究执行机构的科研评价体系政策、面向出版商的研究成果共享政策、面向图书馆的成果整合利用政策等。

(3)与相关政策的协同发展策略。《关于拟定联合国教科文组织开放科学建议书之适宜性的技术、财务和法律问题的初步研究》(*Preliminary Study of the Technical, Financial and Legal Aspects of the Desirability of a UNESCO Recommendation on Open Science*)强调发展开放科学要解决知识产权、数据保护和数据隐私等问题[73]。为此,在研制开放科学政策的同时,需要协调好与知识产权、数据保护、数据隐私等政策的关系。研究需深入探讨版权与使用许可协议、针对敏感人群和数据的保护策略、数据权益政策、数据隐私政策等。

7.2.2 完善全流程的基础设施研究

建设和完善贯穿科研全流程的基础设施是保障开放科学实践顺利开展的基础。纵观目前的基础设施建设,虽然数量众多,但缺乏贯穿科研全流程的集成式基础设施。未来建议建立贯穿研究设计、数据收集、数据清洗、数据分析、成果出版、传播交流、影响评估等科研全流程的科研项目管理平台。美国开放科学框架(Open Science Framework, OSF)、欧洲开放科学云(European Open Science Cloud, EOSC)等基础设施可为研究提供参考。

在建设和完善基础设施的过程中,首先需考虑相关标准。可发现、可访问、可互操作、可重用的 FAIR(Findable, Accessible, Interoperable, Reusable)原则作为一套国际化方法,特别强调增强机器自动查找和使用数据的能力,并支持个人对数据的重用[74],可用于指导基础设施建设。其次要关注基础设施的性能,包括如何建立易于使用的元数据管理系统、如何评估数据质量和来源、如何开发数据集成和统一的分析和可视化功能等。对网站的动态性、数据长期保存、基于不同需求的可参与、可定制、可扩展性等问题也值得探究。最后,还需思考基础设施的交互性,包括平台访问和交互方式、在线科学交流的便捷性、安全性及可持续性等。

7.2.3 加强多方位的主体参与研究

(1)建立健全激励机制。欧盟《开放科学政策平台建议》(*Open Science Policy Platform Recommendations*)[75]提出的关于开放科学的 8 项优先发展建议中,将"奖励和激励"放在首位。有效的激励机制有利于增强研究人员的开放科学意识,促进科学和研究中的开放共享行为。相关研究需要考虑 3 个方面:如何对开放科学工作开展财政支持,如政府部门、资助机构和管理机构如何为开放科学的指导和培训提供经济支持;对参与开放科学实践的研究人员进行认可和奖励,如建立关于数据、代码等更加完善的引用标准,改革职业考核与报酬制度;考虑开放科学及其影响的更广泛的评估指标,如替代计量学指标等。

(2)开展开放科学教育。由欧盟资助的 FOSTER portal 电子学习门户[76]和美国开放科学中心[77]都提供了面向开放科学的各种资料和培训课程,为相关研究提供了实践经验。研究应明确开放科学教育的主体、客体、内容和模式。教育主体以高校和科研院所为主,图书馆也承担着重要角色,需要设计合适的开放科学培训课程和资料。教育客体包括所有研究人员、图书馆员、教师、学生等众多参与到开放科学实践中的利益主体。教育内容包括但不限于开放科学的基本理论、实践方法、道德规范等。教育模式可采用线上与线下相结合、小型课堂与大型讲座并重的多元化培训模式。

(3)推动公众科学发展。由联合国教科文组织发起的 2019 年"争取和平与发展世界科学日"以开放科学为主题,强调开放科学是使所有人都可以访问科学研究和数据的可持续性全球运动[78]。公众科学作为一种新型开放创新模式,有助于推动开放科学发展,集

群体智慧推动科学和社会进步[79]。然而,公众科学也面临着一些问题,如参与者能力差距较大、参与动力不足、经费支撑不够等。未来研究可围绕如何开展公众科学项目、增强公众参与意识展开,具体包括研究公众科学项目的保障制度与运行模式、公众提供数据的质量策略、公众信息素养教育、以信息和通讯技术(Information and Communications Technology,ICT)为代表的技术支撑、图书馆的角色和定位等。

参考文献:

[1] UNESCO. UNESCO mobilizes 122 countries to promote open science and reinforced cooperation in the face of COVID-19 [EB/OL]. [2020－06－10]. https://en.unesco.org/news/unesco-mobilizes-122-countries-promote-open-science-and-reinforced-cooperation-face-covid-19.

[2] UNESCO. UNESCO takes the lead in developing a new global standard-setting instrument on open science[EB/OL]. [2020－06－20]. https://en.unesco.org/news/unesco-takes-lead-developing-new-global-standard-setting-instrument-open-science.

[3] VICENTE-SÁEZ R, MARTÍNEZ-FUENTES C. Open Science now: a systematic literature review for an integrated definition[J]. Journal of business research, 2018, 88: 428－436.

[4] DAI Q, SHIN E, SMITH C. Open and inclusive collaboration in science: a framework[EB/OL]. [2019－12－20]. https://doi.org/10.1787/2dbff737-en.

[5] COOK B G, LLOYD J W, THERRIEN W J. Open science in the field of emotional and behavioral disorders[J]. Education and treatment of children, 2019, 42(4): 579－600.

[6] BANKS G C, FIELD J G, OSWALD F L, et al. Answers to 18 questions about open science practices[J]. Journal of business and psychology, 2019, 34(3): 257－270.

[7] 盛小平,杨智勇. 开放科学、开放共享、开放数据三者关系解析[J]. 图书情报工作,2019,63(17):15－22.

[8] 刘桂锋,钱锦琳,田丽丽. 开放科学:概念辨析、体系解析与理念探析[J]. 图书馆论坛,2018,38(11):1－9.

[9] 黄雨婷,赖彤. 美国开放科学中心:实践进展、特点与启示[J]. 图书与情报,2019(3):105－113.

[10] FECHER B, FRIESIKE S. Open science: one term, five schools of thought[M]//Opening science. New York: Springer, 2014: 17－47.

[11] European Commission. Open science[EB/OL]. [2020－06－15]. https://ec.europa.eu/digital-single-market/open-science.

[12] FOSTER. Open science definition[EB/OL]. [2020－06－02]. https://www.fosteropenscience.eu/taxonomy/term/100.

[13] CHAN L, CUPLINSKAS D, EISEN M, et al. Read the budapest open access initiative[EB/OL]. [2020－06－05]. https://www.budapestopenaccessinitiative.org/read.

[14] UNESCO. Open science movement[EB/OL]. [2020－06－05]. http://www.unesco.org/new/en/communication-and-information/portals-and-platforms/goap/open-science-movement/.

[15] LAAKSO M, WELLING P, BUKVOVA H, et al. The development of open access journal publishing from 1993 to 2009[J]. Plos one, 2011, 6(6): e20961.

[16] PIWOWAR H, PRIEM J, LARIVIÈRE V, et al. The state of OA: a large-scale analysis of the prevalence and impact of open access articles[J]. Peer J, 2018, 6: e4375.

[17] BJÖRK B C. Gold, green, and black open access[J]. Learned publishing, 2017, 30(2):173－175.

[18] VALDERRAMA-ZURIÁN J C, AGUILAR-MOYA R, GORRAIZ J. On the bibliometric nature of a foreseeable relationship: open access and education[J]. Scientometrics, 2019, 120(3): 1031－1057.

[19] ZHANG L, WATSON E. The prevalence of green and grey open access: where do physical science researchers archive their publications? [J]. Scientometrics, 2018, 117(3): 2021－2035.

[20] DOBUSCH L, HEIMSTÄDT M. Predatory publishing in management research: a call for open peer review[J]. Management learning, 2019, 50(5): 607－619.

[21] ROSS-HELLAUER T. What is open peer review? A systematic review[J]. F1000 research, 2017, 6: 1－38.

[22] LAWRENCE S. Free online availability substantially increases a paper's impact[J]. Nature, 2001, 411: 521.

[23] HAJJEM C, HARNAD S, GINGRAS Y. Ten-year cross-disciplinary comparison of the growth of open access and how it increases research citation impact[J]. IEEE data engineering bulletin, 2005, 28(4): 39－47.

[24] MCKIERNAN E C, BOURNE P E, BROWN C T, et al. How open science helps researchers succeed[J]. ELife, 2016, 5: e16800.

[25] HOLLEY R P. Open access: current overview and future prospects [J]. Library trends, 2018, 67(2): 214－240.

[26] WANG X, LIU C, MAO W, et al. The open access advantage considering citation, article usage and social media attention[J]. Scientometrics, 2015, 103(2): 555－564.

[27] TENNANT J P, WALDNER F, JACQUES D C, et al. The academic, economic and societal impacts of open access: an evidence-based review[J]. F1000 research, 2016, 5: 1－49.

[28] FELL M J. The economic impacts of open science: a rapid evidence assessment[J]. Publications, 2019, 7(3): 1－30.

[29] HAMPTON S E, ANDERSON S S, BAGBY S C, et al. The Tao of open science for ecology[J]. Ecosphere, 2015, 6(7): 1－13.

[30] ROUDER J N. The what, why, and how of born-open data[J]. Behavior research methods, 2016, 48(3): 1062－1069.

[31] HARDWICKE T E, MATHUR M B, MACDONALD K, et al. Data availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data policy at the journal cognition [J]. Royal Society open science, 2018, 5(8): 180448.

[32] European Commission. Open science (Open access)[EB/OL].

[2020 – 07 – 20]. https://ec. europa. eu/programmes/horizon 2020/en/h2020-section/open-science-open-access.
[33] ZUIDERWIJK A, JANSSEN M. Open data policies, their implementation and impact: a framework for comparison[J]. Government information quarterly, 2014, 31(1):17 – 29.
[34] NEYLON C. Compliance culture or culture change? The role of funders in improving data management and sharing practice amongst researchers[J]. Research ideas and outcomes, 2017, 3: e14673.
[35] JONES L, GRANT R, HRYNASZKIEWICZ I. Implementing publisher policies that inform, support and encourage authors to share data: two case studies[J]. Insights, 2019, 32(1): 1 – 11.
[36] ROUSI A M, LAAKSO M. Journal research data sharing policies: A study of highly-cited journals in neuroscience, physics, and operations research[J]. Scientometrics, 2020,124: 131 – 152.
[37] NAUGHTON L, KERNOHAN D. Making sense of journal research data policies[J]. Insights, 2016, 29(1): 84 – 89.
[38] GORMAN D M. Availability of research data in high-impact addiction journals with data sharing policies[J]. Science and engineering ethics, 2020, 26(3): 1625 – 1632.
[39] Data Policy and Standardisation and Implementation Interest Group. Journal and publisher research data policy master framework DRAFT v1.2 Feb 2018[EB/OL]. [2020 – 06 – 20]. https://docs. google. com/document/d/1DTAfOKkE1a2n2f_1hGcrXlrw-5Tq_AL5tk-ju8B82_E/edit.
[40] YANCY C W, Harrington R A, Bonow R O. Data sharing-The time has (not yet?) come[J]. JAMA cardiology, 2018, 3(9): 797 – 798.
[41] ABELE-BREHM A E, GOLLWITZER M, STEINBERG U, et al. Attitudes toward open science and public data sharing: a survey among members of the German Psychological Society[J]. Social psychology, 2019, 50(4): 252 – 260.
[42] ROCKHOLD F W. Statistical controversies in clinical research: data access and sharing-can we be more transparent about clinical research? Let's do what's right for patients[J]. Annals of oncology, 2017, 28(8): 1734 – 1737.
[43] THOROGOOD A. Canada: will privacy rules continue to favour open science? [J]. Human genetics, 2018, 137(8): 595 – 602.
[44] HOEYER K, TUPASELA A, RASMUSSEN M B. Ethics policies and ethics work in cross-national genetic research and data sharing: flows, nonflows, and overflows[J]. Science, technology & human values, 2017, 42(3): 381 – 404.
[45] ZUIDERWIJK A, SPIERS H. Sharing and re-using open data: a case study of motivations in astrophysics[J]. International journal of information management, 2019, 49: 228 – 241.
[46] ROCKHOLD F, BROMLEY C, WAGNER E K, et al. Open science: the open clinical trials data journey[J]. Clinical trials, 2019, 16(5): 539 – 546.
[47] CHAWINGA W D, ZINN S. Global perspectives of research data sharing: a systematic literature review[J]. Library & information science research, 2019, 41(2): 109 – 122.
[48] ROSS J S, WALDSTREICHER J, BAMFORD S, et al. Overview and experience of the YODA Project with clinical trial data sharing after 5 years[J]. Scientific data, 2018, 5: 180268.
[49] LEFAIVRE S C, BEHAN B, VACCARINO A L, et al. Big data needs big governance: best practices from Brain-CODE, the Ontario Brain Institute's neuroinformatics platform[J]. Frontiers in genetics, 2019, 10: 191.
[50] CAPUZZI S J, KIM I S J, LAM W I, et al. Chembench: A publicly accessible, integrated cheminformatics portal[J]. Journal of chemical information and modeling, 2017, 57(2): 105 – 108.
[51] National Academies of Sciences, Engineering, and Medicine. Reproducibility and replicability in science[M]. Washington: The National Academies Press, 2019.
[52] MARWICK B. Computational reproducibility in archaeological research: basic principles and a case study of their implementation [J]. Journal of archaeological method and theory, 2017, 24(2): 424 – 450.
[53] BRITO J J, LI J, MOORE J H, et al. Recommendations to enhance rigor and reproducibility in biomedical research[J]. Gigascience, 2020, 9(6): giaa103.
[54] HETTNE K, PROPPERT R, NAB L, et al. ReprohackNL 2019: how libraries can promote research reproducibility through community engagement[EB/OL]. [2020 – 07 – 17]. https://doi. org/10.31235/osf. io/6f4zv.
[55] MELLOR D T. Guidelines for Transparency and Openness Promotion (TOP) in journal policies and practices "The TOP Guidelines"[EB/OL]. [2020 – 07 – 17]. https://osf. io/9f6gx/wiki/Guidelines/.
[56] LEWANDOWSKY S, OBERAUER K. Low replicability can support robust and efficient science[J]. Nature communications, 2020, 11: 358.
[57] LARSON M J. Improving the rigor and replicability of applied psychophysiology research: sample size, standardization, transparency, and preregistration[J]. Biofeedback, 2020, 48(1): 2 – 6.
[58] BENTWICH M. Changing the rules of the game: addressing the conflict between free access to scientific discovery and intellectual property rights[J]. Nature biotechnology, 2010, 28(2): 137 – 140.
[59] GANS J S, MURRAY F E, STERN S. Contracting over the disclosure of scientific knowledge: intellectual property and academic publication[J]. Research policy, 2017, 46(4): 820 – 835.
[60] OWENS B. Montreal instiute going "open" to accelerate science [EB/OL]. [2020 – 07 – 10]. http://www. sciencemag. org/news/2016/01/montreal-institute-going-open-accelerate-science.
[61] Open Data Institute. Publisher's guide to open data licensing[EB/OL]. [2020 – 07 – 10]. https://theodi. org/article/publishers-guide-to-open-data-licensing/.
[62] KORN N, OPPENHEIM C. Licensing open data: a practical guide

[EB/OL]. [2020－07－10]. http://discovery. ac. uk/files/pdf/Licensing_Open_Data_A_Practical_Guide. pdf.
[63] ROSEN L. Open source licensing: software freedom and intellectual property law[M]. Upper Saddle River: Prentice Hall, 2004.
[64] Open Source Hardware Association. Open-Source hardware FAQ [EB/OL]. [2020－07－15]. http://www. oshwa. org/faq/.
[65] LUIS FELIPE R M, KAUTTU P, LAIA P P, et al. Open hardware licences: Parallels and contrasts[EB/OL]. [2020－08－31]. https://helda. helsinki. fi/bitstream/handle/10138/312235/ki0119833enn. pdf? sequence=1.
[66] ALTUNAY M, AVERY P, BLACKBURN K, et al. A science driven production cyberinfrastructure-the open science grid[J]. Journal of grid computing, 2011, 9(2): 201－218.
[67] DE ROURE D, GOBLE C, ALEKSEJEVS S, et al. Towards open science: the myExperiment approach[J]. Concurrency and computation: practice and experience, 2010, 22(17): 2335－2353.
[68] NISHIKAWA K. How are research data governed at Japanese repositories? A knowledge commons perspective[J]. Aslib journal of information management, 2020, 72(5): 837－852.
[69] XAFIS V, LABUDE M K. Openness in big data and data repositories: the application of an ethics framework for big data in health and research[J]. Asian bioethics review, 2019, 11(3): 255－273.
[70] OECD. Business models for sustainable research data repositories [EB/OL]. [2020－07－22]. https://doi. org/10. 1787/302b12bb-en.
[71] United Nations. Open science conference 2019[EB/OL]. [2020－07－10]. http://research. un. org/conferences/OpenScienceUN.
[72] 顾立平. 科研模式变革中的数据管理服务:实现开放获取、开放数据、开放科学的途径[J]. 中国图书馆学报,2018,44(6):43－58.
[73] UNESCO. Preliminary study of the technical, financial and legal aspects of the desirability of a UNESCO recommendation on open science[R/OL]. [2020－07－05]. https://unesdoc. unesco. org/ark:/48223/pf0000370291.
[74] WILKINSON M, DUMONTIER M, AALBERSBERG I. et al. The FAIR guiding principles for scientific data management and stewardship[J]. Scientific data, 2016, 3: 160018.
[75] European Commission. Open science policy platform recommendations[EB/OL]. [2020－07－02]. https://ec. europa. eu/research/openscience/pdf/integrated_advice_opspp_recommendations. pdf#view=fit&pagemode=none.
[76] FOSTER. About FOSTER[EB/OL]. [2020－07－02]. https://www. fosteropenscience. eu/about.
[77] Center for Open Science. Openness and reproducibility research practices training[EB/OL]. [2020－07－02]. https://cos. io/our-services/training-services/.
[78] UNESCO. World science day for peace and development[EB/OL]. [2020－07－02]. https://zh. unesco. org/commemorations/worldscienceday#theme.
[79] 金瑛,张晓林,胡智慧. 公众科学的发展与挑战[J]. 图书情报工作, 2019,63(13):28－33.

作者贡献说明:
黄如花:选题,指导撰写论文;
赵洋:收集数据,撰写并修改论文;
黄雨婷:提出论文修改意见,校稿。

Research Progress in the International Research of Open Science

Huang Ruhua　Zhao Yang　Huang Yuting
School of Information Management, Wuhan University, Wuhan 430072

Abstract: [**Purpose/significance**] Open science can make scientific research more collaborative, transparent and efficient, which has attracted widespread attention worldwide. Sorting out and analyzing the current research themes of open science from an international perspective is conducive to promoting the development of this field and providing reference and experience for subsequent research. [**Method/process**] Using a mixed method which combine literature survey and information visualization, this paper analyzed articles on open science included in the Web of Science platform, combined with relevant policies, research reports, news and other literature. The theme distribution of international research on open science was analyzed. [**Result/conclusion**] The research topics of open science include: open access, data sharing, research reuse, knowledge innovation, and infrastructure construction. In the future, it is recommended to conduct a systematic and in-depth study from three aspects of policy system, infrastructure and subject participation.

Keywords: open science　open data　open access　data sharing